DOCUMENT RESUME

ED 050 172 TM 000 568

AUTHOR Porter, Andrew C.
TITLE How Errors of Measurement Affect ANOVA, Regression Analyses, ANCOVA and Factor Analyses.
PUB DATE Jan 71
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Analysis of Covariance, *Analysis of Variance, Correlation, *Factor Analysis, Measurement, Models, *Multiple Regression Analysis, *Standard Error of Measurement, Statistical Analysis, Statistics

ABSTRACT

In this paper problems caused by the existence of errors of measurement are identified for factor analysis, regression analysis, ANOVA, and ANCOVA. At least one detrimental effect is shown to exist for each type of analysis. When a researcher's interest is with infallible variables, he runs the risk of biased results from all of the procedures except ANOVA. The estimates of parameters in all four procedures suffer from inflated error variance. Some partial solutions are indicated, but clearly more work is needed on several of the problems. Most statistical procedures have been developed for models where variables are assumed to be free from errors of measurement. Since almost all educational research involves use of fallible variables, it is important that the effects of errors of measurement for the various models be understood and that the understanding be reflected in current research practice. (Author/GS)

ED050172

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

HOW ERRORS OF MEASUREMENT AFFECT ANOVA, REGRESSION ANALYSES, ANCOVA AND FACTOR ANALYSES

Paper presented at the 1971 AERA convention

Andrew C. Porter

Michigan State University

January 1971

TM 000 568

The other papers in the symposium have concentrated on the effects of errors of measurement on various types of correlation measures. In contrast, the purpose of the present paper is to consider the properties of ANOVA, regression analyses, ANCOVA, and factor analyses when employed on fallible variables, i.e., variables which contain errors of measurement. Because of the broad range of topics to be considered, the discussion of each will be somewhat brief. For each topic the major problems of concern to the user of statistics will be identified; when available, solutions to the problems will be indicated; and several of the more important references will be cited.

Factor analysis is considered first because the primary concern is with the properties of the elements in the correlation matrix to be factored, and so perhaps will facilitate a transfer from the preceeding papers. The topic of regression analysis is considered second because it logically leads the way to subsequent discussions of ANOVA and ANCOVA. As Fisher (1932) has said, ANCOVA "combines the advantages and reconciles the requirements of the two very widely applicable procedures known as regression and analysis of variance."

Factor Analysis

Although the literature abounds with articles on the theory and use of factor analysis, few have been concerned with the nature of the correlations in the matrix to be factored. When discussing the nature of correlation coefficients in a matrix to be factored, it is useful to make a distinction between manifest and latent relationships. Manifest relationships are those

obtained from variables as they are observed, while latent relationships are those which "may be inferred to exist between variables and which are masked or distorted by various kinds of errors and constraints" (Carroll, 1961, p. 351). When used as an instrument to facilitate theory building, factor analysis should operate on a matrix of correlation coefficients which reflect the latent relationships among the variables.

Carroll (1961) has identified the following four categories of errors and constraints which can affect the value of a Pearson Product-Moment correlation coefficient:

1. errors of scaling,
2. errors of scale-dependent selection,
3. scedastic errors of measurement,
4. topastic errors of measurement.

Errors of scaling result when a dichotomy is forced on a continuous variable. An error of scale-dependent selection occurs when the sample is taken such that subjects with extreme scores on either or both variables are not selected. Scedastic errors of measurement are those dealt with in classical measurement theory, i.e. independent of each other and the latent variable with expected value zero; and topastic errors of measurement are created when a subject guesses correctly on a multiple choice item.

Carroll states that except for scedastic errors of measurement the errors and constraints can alter the rank of a correlation matrix and cause subsequent factor analysis to yield spurious results. The exception made for scedastic errors probably stems from results given by Roff (1937) and later supported by Saunders (1948) which indicate that

1. the rank of the correlation matrix, R, is unaffected by scedastic errors;

2. the communality of an infallible variable is equal to the communality of the fallible variable divided by the reliability of the fallible variable;

and 3. the factor pattern matrix of the infallible variables, F^*, is related to the factor pattern matrix of the fallible variables, F, by the equation

$$F^* = AF,$$

where A is a diagonal matrix of the inverses of the square roots of the reliabilities of the fallible variables.

As will be pointed out later, the exception made for scedastic errors was a mistake since the relationships given by Roff are based on the seldom met assumption that communalities are known rather than estimated.

A matrix of Pearson Product-Moment coefficients for variables containing errors of scaling generally results in a factor analysis solution which contains a difficulty factor, although Kaiser (1970) has recently indicated that Guttman's image analysis may not. Carroll (1961) recommends the use of tetrachoric coefficients to side step the problem of errors of scaling. The problems of errors caused by scale-dependent selection have not been solved except by the obvious method of avoiding them via careful sampling procedures. The procedures of correcting correlations for restriction of range (Guilford, 1959; Bryant, 1970) might be useful, but to my knowledge the consequence of using such corrected coefficients in factor analysis have not been investigated. Carroll (1961) has given a method of correcting joint and marginal distributions of the observed variables for topastic error, and suggests that correlations be calculated on the corrected distributions.

Glass (1966) has indicated the problems with several methods of factor analysis which result from scedastic errors. In particular he has shown that a components analysis of the correlation matrix with ones in the main diagonal and off diagonal elements corrected for attenuation is not necessarily of the

same order as components analysis of the uncorrected matrix. Factoring the correlation matrix with Guttman's lower bound to communalities in the main diagonal does not result in the relation $F^* = AF$, nor does Rao's canonical factor analysis. On the positive side, Glass demonstrated that $F^* = AF$ for Kaiser's alpha factor analysis.

Briefly, alpha factor analysis is an iterative procedure which starts with a principal axis factorization of the matrix

$$H_1^{-1} (R - I)H_1^{-1} + I,$$

where H_1 is a diagonal matrix of the multiple correlation of each variable with the remaining variables and I is the identity matrix. As many factors are retained as there are latent roots of the above matrix which exceed one. The solution is used to calculate new estimates of the communalities which are used to replace those in H_1 to yield H_2. The procedure is repeated until the estimates of the communalities have converged according to an arbitrary apriori criterion. Glass has shown that by replacing H_2^{-1} with $H_3^{-1} = H_2^{-1} A$, alpha factor analysis results in the relationship $F^* = AF$ provided that H_3 and H_1 will iterate to the same matrix of communalities. A necessary assumption is that the elements of A satisfy $h_j^2 \leq 1/a_j^2 \leq 1$ for all j, where h_j^2 is the communality of the j <u>th</u> variable. Glass has demonstrated on several well known examples that the two matrices do converge on the same parameters. Glass (1966, p. 559) further derived that "Alpha factor analysis applied to fallible and infallible variables separately will be equivalent in terms of number of factors, complexities of corresponding variables, and patterns of simple structure"; "Normal varimax rotations of F and F* will yield derived solutions F_1 and F_1^* such that $AF_1 = F_1^*$; and "The alpha factor scores for the fallible and corresponding infallible variables may be considered identical."

The above suggest that a researcher interested in using factor analysis to investiage the latent structure of variables should first be careful in his sampling of subjects to avoid possible problems brought on by errors of scale-dependent selection. Second, he should employ Carroll's correction to the marginal and joint distributions of each pair of variables to control for topastic errors. Third, he should calculate tetrachoric coefficients on the corrected distributions to avoid the problem of difficulty factors arising from errors of scaling. Finally, given the resulting correlation matrix he should use Kaiser's alpha factor analysis to side step the problem of scedastic errors of measurement. The use of alpha factor analysis seems to follow Kaiser's recently stated first principle in dealing with problems of factor analysis, i.e. "It don't make no never-mind." What is meant by the principle he says is "that when faced with a crucial decision, don't try to settle it; rather, avoid it!" (Kaiser, 1970, p. 403).

Regression Analyses

Madansky (1959) identifies three basic types of regression relationships which are generally referred to, although not always (Lindley, 1947), as regression, structural, and functional. Regression is defined as the appropriate relationship for predicting one set of scores from another. Because the purpose is to predict one set of scores on the basis of another set of scores, the relation should be defined by the observations, making the least-squares estimate appropriate, i.e., the manifest relationship is of interest. A structural relationship is defined by the true parts of the variables, when the independent variable is random. A functional relationship is also defined by the true parts of the variables, but the independent variable is fixed, and the true variables are perfectly correlated. These

last two types of relationships fall under the general category of latent relationships and are of interest when theory building or testing is the objective. Because problems of prediction and the regression relationship are familiar, and because functional relationships are probably quite rare in educational research, the following discussion emphasizes structural relationships. Both Madansky (1959) and Cochran (1968) offer excellent reviews of the work done on estimating structural relationships.

First mention of the inappropriateness of a least-squares estimate of the structural relation, when the variables are fallible, was made by R. J. Adcock (1878). However, in an early review of the problem of estimating structural relations, Roos (1937, P. 7) credits Corrado Gini in 1921 as the first to recognize that "if the errors of X and Y are independent, then the least-squares β is larger than the β of the actual line of best fit." Since then, a considerable body of literature has dealt with the problem of estimating the structural relation when both variables are fallible.

In the first half of a paper by Berkson (1950), the problem of estimating a structural relation is stated, and an analytic demonstration of the bias of the least-squares solution is given for the case of scedastic errors of measurement. Berkson's demonstration shows that the β defined by a fallible dependent variable, Y, and a fallible independent variable, X, is equal to the β defined by the true parts, multiplied by the ratio of the variance of the true parts of X over the variance of the true parts of X plus the variance of the error parts of X. In the notation adopted here

$$\beta_{Y.X} = \frac{\sigma^2_T}{\sigma^2_T + \sigma^2_E} \beta_{Y.T} ,$$

where T denotes the true parts of X. The ratio which defines the bias of the least-squares β is equivalent to the measurement theory definition of the reliability of X (Gulliksen, 1950, p. 25). Thus

$$\beta_{Y.X} = \rho_{XX}\beta_{Y.T} ,$$

where ρ_{XX} denotes the reliability of X. It should be noted at this point that the bias does not depend on the dependent variable, Y, and therefore, the fallibility of Y does not affect the least-squares estimate of the structural relation. Rather than offering Berkson's derivation, a derivation which is consistent with measurement theory seemed more appropriate.

Let ρ_{YT} be the correlation of Y and T, and ρ_{XY} be the correlation of Y and X.

$$\beta_{Y.X} = \rho_{YX} \frac{\sigma_Y}{\sigma_X} ,$$

but $\rho_{YX} = \sqrt{\rho_{XX}}\ \rho_{YT}$, (Gulliksen, 1950, p. 105)

and $\sigma_X = \frac{\sigma_T}{\sqrt{\rho_{XX}}}$

Therefore, $\beta_{Y.X} = \rho_{YT}\ \rho_{XX} \frac{\sigma_Y}{\sigma_T}$

and $\beta_{Y.X} = \rho_{XX}\ \beta_{Y.T}$.

Because the expected values of the fallible variables are equal to the expected values of the infallible variables

$$\alpha_{Y.X} = \alpha_{Y.T} + (\beta_{Y.T} - \beta_{Y.X})\ E(X)$$

where α defines the Y intercept of the regression line.

Under the same classical model for errors of measurement but in different notation, Cochran (1970) gives the relationship between the manifest and structural multiple regression equations as

$$\beta_{Y.X_i} = \beta_{Y.T_i} - \sum_{j=1}^{k} (1-\rho_{X_jX_j})\ \beta_{Y.T_j} ,$$

where i=1, ..., k denotes which independent variable. The relationship between the manifest and structural intercepts is

$$\alpha' = \alpha + \sum_{j=1}^{k} (\beta_{Y \cdot T_j} - \beta_{Y \cdot X_j}) E(X_j),$$

where α' denotes the manifest intercept and α denotes the structural intercept. For k=1 the above relationships reduce to those given earlier for a single independent variable. Cochran concludes that the direct effect of errors in an independent variable is to decrease the absolute value of its associated β weight by at least a factor of its reliability coefficient. Cochran (1968) has also given a statement of the above relationships under a slightly less restricted model for errors of measurement. Lindley (1947) has demonstrated that even when the structural relationship is linear, the manifest relationship need not be linear. If the infallible variables are multivariate normal then the errors of measurement must also be multivariate normal to assure a linear manifest relationship.

Karl Pearson (1901) offered the first approach to gain any prominence for estimating structural relations, using fallible variables. Pearson proposed minimizing the sum of the squared normal deviates of the observed points from a line which has come to be called the orthogonal regression line. Allen (1939) has shown that an orthogonal regression line is dependent upon the choice of units used in measuring the variables. The orthogonal regression line can always be the structural relation, if the units chosen to measure the variables happen to be the right ones. There is no practical way of determining what the units should be without additional information (Allen, 1939, p. 198). C. F. Roos (1937, p. 18) offers an orthogonal regression line solution, which is invariant to the metric of the variables but requires <u>apriori</u> information about the magnitude of the errors.

Most of the developments since Pearson can be placed in one of the four categories: the method of grouping, the use of instrumental variables, the use of variance components, and the Berkson case. To supply detailed information about the various types of estimates would not be appropriate here, but a brief acquaintance with one of the more popular techniques in each category should provide helpful background.

Wald (1940) provides a method for estimating structural relations, based on the method of grouping. Letting X denote the independent variable and Y the dependent variable, Wald makes the following assumptions:

1. Errors on X are uncorrelated and have a common distribution,
2. Errors on Y are uncorrelated and have a common distribution,
3. Errors on X are uncorrelated with errors on Y,
4. There is a single linear relation between the true variables,
5. Observations on X, can be divided into subgroups in such a way that the true part of any observation will be in the same subgroup as its associated observed score.

The location of the line, identifying the structural relationship, is estimated by the mean coordinates. To estimate slope, Wald first orders the observations on X and divides them into two groups of equal size. The join of the mean coordinates of the two groups provides the estimate of slope. Wald derived the variance of the above estimate of a structural relation and provided a test of hypotheses about the size of the slope. Wald demonstrated that his estimate of structural relation is a consistent estimate and that the hypothesis test is exactly correct if the assumptions are satisfied.

Bartlett (1949) extended Wald's estimate by proposing that the subgroups for estimating slope be the upper and lower one-third of the ordered

X values. Bartlett provides confidence interval estimates for his estimate of slope and shows that, when X has a rectangular distribution, the estimate is more efficient than Wald's. More generally, the efficiency of a criterion for grouping is dependent upon the distribution of the independent variable. For a normally distributed independent variable, Madansky (1959) has shown that the most efficient criterion for grouping is to use the upper-lower twenty-seven percent. Madansky (1959, p. 184) offers a table, which indicates the most efficient criterion for grouping several different types of distributions. Besides being dependent upon the distribution of the independent variable, the method of grouping further requires knowledge that the errors of the independent variable are independent of the grouping. Without this additional information, the estimates are not necessarily consistent (Neyman, 1951).

Reiersol (1945) originated the strategy of using instrumental variables in the estimation of structural relations and the approach has since been developed by others. An approach, taken by Durbin (1954), represents the simplest use of an instrumental variable, although probably the least often applicable. The assumptions are the same as the first four given, in regard to the method of grouping. Let z be an infallible instrumental variable which is correlated with the true parts of the independent variable but not with the error parts. Then

$$b = \frac{\sum_{i=1}^{N} z_i Y_i}{\sum_{i=1}^{N} z_i X_i}$$

is a consistent estimate of the structural relation of Y on X. Durbin offers a proof of b's consistency and derives a confidence region for the parameter

of the structural relation by making use of the fact that true regression $Y - \beta_X$ on z is zero. As would be expected, Durbin's estimate is less efficient than a least-squares estimate, based on the unobservable true parts of the fallible variables.

Reiersol (1945) has considered the estimation of structural relations, using fallible instrumental variables. Reiersol's estimate requires observations on two instrumental variables, both of which are correlated with the true parts of the independent variable, but not with the error parts, and further states that these instrumental variables have some known linear relation.

Geary (1949, p. 30) states that the accuracy of using instrumental variables to estimate structural relations is dependent upon the correlation of the instrumental variables with the dependent and independent variables. Even if instrumental variables can be identified as having the required properties, which in itself seems unlikely, their use represents additional cost. Madansky (1959, p. 188) identifies Durbin's use of an infallible instrumental variable as equivalent to the grouping method, when the instrumental variable is restricted to the values -1, 0, and +1.

The variance components procedure for estimating structural relations, using fallible variables, began with Tukey (1951), and requires additional information, similar to that necessary for grouping. The data must be in the form of n_i observations, X_{ij}, on each of N X_i's, i.e., two or more groups must be identified into which the data may be divided. The procedure is to do one-way analyses of variance on the variables X, Y, and XY, as in an analysis of covariance. The mean squares and expected mean squares of the computations are given in Table 1. (Madansky, 1959, p. 189).

An inspection of the expected mean squares in Table I indicates

$$b=\sqrt{\frac{III - VI}{I - IV}}$$

is an estimate of the structural relation of Y on X. Tukey shows the estimate to be consistent as $N \to \infty$ and some $n_i \to \infty$, and provides confidence intervals for β, both in the case of a functional relation and in the case of a structural relation.

Berkson's method (1950) for estimating structural relations is dependent upon his sampling model which distinguishes between a "controlled observation" and an "uncontrolled observation." Let $x = X + u$ be an observation on the fallible independent variable, broken down into a true part X and an error part u; and similarly, let $y = Y + v$ be a fallible observation on the dependent variable. An "uncontrolled observation" on the fallible independent variable is defined as X being fixed, and u being a random variable, independent of X. The structural relation is $Y = \alpha + \beta X$ and by substitution $y = \alpha + \beta x + (v - \beta u)$. The above equation is not a standard regression model because the random error $(v - \beta u)$ is not independent of x. The purpose of taking a "controlled observation" is not to estimate X, but rather, to bring the observed quantity to a set value. The attempt is to get x each time an observation is made, but because of errors of measurement, the observation represents $x - u$. The sampling model causes X to vary while x remains fixed; therefore, u is independent of x. Since for a "controlled observation" the error of measurement is independent of the observation, the least-squares estimate of slope is also an unbiased estimate of the structural relation. A fixed x also eliminates the population correlation of x and y, which means the regression line y on x is the same as the regression line x on y. Scheffé (1958) offers a slight

TABLE 1

Tukey's Variance Components*

Source		Mean Square	Expected Mean Square
Between Sets	I	$\sum_{i=1}^{N} n_i(\bar{X}_i\cdot-\bar{X}_{..})^2 / (N-1)$	$\sigma_u^2 + [N^2 - \sum_{i=1}^{N} n_i^2) / (Nn-n)]\ \sigma_x^2$
	II	$\sum_{i=1}^{N} n_i(\bar{X}_{i.}-\bar{X}_{..})(\bar{Y}_{i.}-\bar{Y}_{..})$ /(N-1)	$\mathrm{cov}(u,v) + [(N^2 - \sum_{i=1}^{N} n_i^2) /$ (Nn-n)] $\beta\sigma_x^2$
	III	$\sum_{i=1}^{N} n_i(\bar{Y}_i - \bar{Y}_{..}) / (N-1)$	$\sigma_v^2 + [(N^2 - \sum_{i=1}^{N} n_i^2)/(Nn-n)]\beta^2\sigma_x^2$
Within Sets	IV	$\sum_{i=1}^{N}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2 / (N-1)$	σ_u^2
	V	$\sum_{i=1}^{N}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})(Y_{ij}-\bar{Y}_{i.})$ /(N-1)	$\mathrm{cov}(u,v)$
	VI	$\sum_{i=1}^{N}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2 / (N-1)$	σ_v^2

*u = error part of X

v = error part of Y

modification of Berkson's model by allowing u and v to be independent variables, with variances σ_X^2 and σ_Y^2. The modification is dependent upon the manner in which replication is achieved. Scheffé supposes that replicates are taken by changing the controlled bariable from its previous value and bringing it back for each replicate; whereas Berkson assumes the controlled variable remains unchanged. Scheffé develops confidence intervals for the slope and intercept, under the assumptions of his restatement of the Berkson case.

ANOVA and ANCOVA

The effects of scedastic errors of measurement in ANOVA are the same as they are in ANCOVA with a fallible dependent variable and an infallible covariable. Because scedastic errors have an expected value of zero and are independent of the true parts of the variable, the least squares procedures of ANOVA provide unbiased estimates of the parameters in its linear model. In ANCOVA the least squares estimate of the slope of the structural relation of the dependent variable on the covariable is unbiased if the covariable is infallible, as seen earlier, and so the least squares procedures of ANCOVA also provide unbiased estimates of the parameters in its linear model. Cochran (1968) has considered the problem of estimating the parameters in the linear model of ANOVA for a less restrictive model of errors of measurement and comes to the same favorable conclusion.

Given the usual assumptions, the F test statistics for ANOVA on a fallible dependent variable and the F test statistics of ANCOVA on a fallible dependent variable and an infallible covariable will follow theoretical F distributions. The only detrimental effect that scedastic errors of measurement in the dependent variable have is to decrease precision and thus increase

the probability of a Type II error. For ANOVA the decrease in precision is seen by the equation

$$\sigma_Y^2 = \sigma_T^2 + \sigma_e^2 ,$$

where σ_Y^2 denotes the variance of the fallible dependent variable, σ_T^2 is the variance of the infallible dependent variable which is inflated by σ_e^2 the variance of the errors of measurement. The error variance of a one-way ANCOVA is

$$\sigma_Y^2(1 - \rho_{XY}^2)\ [1 + \frac{1}{f_e - 2}],$$

where σ_Y^2 is the variance of the fallible dependent variable, ρ_{XY} is the correlation of the fallible dependent variable with the infallible covariable and f_e denotes the degrees of freedom for estimating error variance. The σ_Y^2 term is inflated by errors of measurement in the same way as shown above for ANOVA. Further, ρ_{XY} is attenuated by errors of measurement in Y. Both effects cause a loss in precision due to a fallible dependent variable. Sutcliffe (1958) has derived the expected mean squares for a one-way, fixed effects ANOVA for a fallible dependent variable. By using his table of expected mean squares, Sutcliffe points out the increase in the probability of a Type II error caused by the errors of measurement. Box (1961) has considered whether the ANOVA least squares methods employed on fallible variables provide unbiased estimates of error variances in factorial and fractional factorial designs. He concludes that the estimates are unbiased for both high order interactions and the pooled variance of replications within cells.

The use of a fallible covariable in ANCOVA can cause a far more distressing problem than decreased precision. First, consider the linear model

of a one-way fixed-effects ANCOVA with a single random covariable. Note that a random covariable represents a relaxing of the restriction in the classical model that covariables be fixed. DeGracie (1968) has shown that the usual ANCOVA procedures applied to data with a random covariable still provide unbiased estimates of the parameters in the linear model as well as valid test statistics. The only difference from classical results is that the variances of the various estimates are averaged across all values of the covariable. A random covariable is certainly more representative of practice in educational research than is a fixed covariable. The linear model is

$$Y_{ij} = \mu_{Y..} + \alpha_{.j} + \beta_{Y.X}(X_{ij} - \mu_{X..}) + e_{ij},$$

where Y_{ij} and X_{ij} are the i th observations in the j th treatment for the dependent and covariable respectively,

$\mu_{Y..}$ is the constant for true mean response,

$\alpha_{.j} = \mu_{Y_{.j}} - \mu_{Y..} - \beta_{Y.X}(\mu_{X.j} - \mu_{X..})$ is the j th treatment effect,

$\beta_{Y.X}$ is the common within treatment slope of the regression of Y on X,

and e_{ij} are random variables assumed to be normally distributed independent of each other with zero mean and common variance.

When ANCOVA is employed as a method for gaining precision in designs where experimental units are randomly assigned to treatments and where the covariable is observed antecedent to the experiment, $\mu_{X_{.j}} = \mu_{X_{.j'}}$ for all $j \neq j'$, and the treatment effects, $\alpha_{.j}$, reduce to those for ANOVA, i.e. $\mu_{Y_{.j}} - \mu_{Y..}$. The only negative effect of a fallible covariable beyond that of a fallible dependent variable is a further attenuation of ρ_{XY} causing a loss in precision as seen earlier. However, when ANCOVA is used in attempt to control for systematic initial between group differences on the covariable,

in general $\mu_{X_{.j}} \neq \mu_{X_{.j'}}$ and the treatment effects will reflect the nature of the correction term

$$\beta_{Y \cdot X} (\mu_{X_{.j}} - \mu_{X..}).$$

This second use of ANCOVA has become common practice in educational research since suggested by Campbell and Stanley (1963) for use on their quasi-experiments. Lord (1960), Smith (1957) and Thorndike (1942) as well as several others more recently have recognized that even though X and Y are fallible, the regression line relevant to ANCOVA is the structural relationship. Since

$$\beta_{Y \cdot X} = \rho_{XX} \beta_{Y \cdot T},$$

the usual ANCOVA procedures provide and test biased estimates of the treatment effects when $\mu_{X_{.j}} \neq \mu_{X_{.j'}}$. Further, by plugging in different possible values of ρ_{XX}, $\beta_{Y \cdot T}$, and $\mu_{.j}$ it is seen that usual ANCOVA procedures can provide non zero treatment effects when the actual effects are zero or zero treatment effects when the actual treatment effects are non zero. (Porter, 1967) Lord (1960) offers a graphic demonstration of the same problem.

Lord (1960) was the first to provide a statistical procedure that yields and tests unbiased estimates of the correct treatment effects when analyzing data from quasi experiments having a fallible dependent variable and a fallible co-variable. His test statistic is asymptotically distributed normal and is limited to consideration of only two levels of the independent variable. The necessary data are observations on the dependent variable and duplicate observations on the covariable, where the duplicate measures follow the test-retest paradigm of classical measurement theory.

I have developed another approach to the problem of obtaining and testing unbiased estimates of the correct treatment effects when analyzing

data from quasi experiments having a fallible dependent variable and a fallible covariable. At least computationally my procedure can be used in any design where classical ANCOVA can be used (Porter, 1967 and 1968). Essentially the procedure I have suggested and investigated substitutes an estimated true score covariable for the observed fallible covariable and then employs classical ANCOVA procedures. The estimated true score variable in a one way ANCOVA is defined as

$$\hat{T}_{ij} = \bar{X}_{.j} + \rho_{XX} (X_{ij} - \bar{X}_{.j}) ,$$

For more complex designs the estimated true score covariable would follow the same form except that the observations would be deviated from the respective cell means. The important properties of an estimated true score covariable are that it is a linear transformation of the fallible covariable and

1) has the same treatment group and grand means as the fallible and unobserved infallible covariable,

2) has the same correlation with the dependent variable as does the fallible covariable,

and 3) the slope, $\beta_{Y.\hat{T}}$, is equal to the desired slope of the structural relation, $\beta_{Y.T}$.

From the above three points it follows that use of classical ANCOVA on the fallible dependent variable and the estimated true score covariable will provide unbiased estimates of the treatment main effects

$$\alpha_{.j} = \mu_{Y_{.j}} - \mu_{Y..} - \beta_{Y.T}(\mu_{X_{.j}} - \mu_{X..}) ,$$

and that the F test statistic will follow the theoretical F distribution given the usual assumptions of ANCOVA plus the assumption that the reliability of X is common across all treatment groups. When the reliability of X is

not known, an estimate must be used which introduces an additional source of variation into the model. With the additional source of variation due to estimating ρ_{XX}, ANCOVA using estimated true scores as the covariable no longer conforms to the classical model and the distribution of the F test statistic must be questioned. Also of interest is a comparison of the small sample properties of Lord's procedure to the one I have proposed when there are only two levels of the treatment independent variable.

In a recent study (Porter, 1967) I used the Monte Carlo approach to investigate the effects of sample size, the reliability of the covariable, and the correlation of the dependent variable with the covariable on the small sample properties of Lord's statistic and on the distribution of the F statistic calculated from ANCOVA using estimated true scores as the covariable. The ANCOVA procedure was investigated for both two and four levels of the independent variable. Each empirical distribution was based on 1000 values of the test statistic and the method for estimating reliability conformed to the test retest paradigm suggested by Lord. Further, the covariable was random.

The results indicated that when the reliability of the covariable was as low as .5, the distribution of Lord's statistic was a very poor approximation of the normal distribution. As the reliability of the covariable increased, the number of observations per treatment group necessary for a good approximation became less. The size of the correlation of the covariable with the dependent variable had an inverse effect on the rapidity of convergence. Samples of size 20 or greater per treatment group seemed to provide for sufficient convergence of Lord's statistic for intermediate values of reliability and correlation. A sample size, larger than 20, was necessary in order for the distribution of Lord's statistic to converge upon the normal when reliability, correlation, or both were low.

The analysis of covariance, using estimated true scores as a covariable, also required samples of size 20 or larger per treatment group for the theoretical F to serve as a useful reference distribution. For covariables having reliability of .7 or .9, the generated distributions of F based on analysis of covariance using estimated true scores as the covariable were in close agreement with the corresponding theoretical distributions of F. When the reliability of the covariable was .5, the agreement of the generated distributions of F to their theoretical counterparts were not quite as good, but still close. The degree of agreement suffered a greater decrease caused by a decrease in the reliability of the covariable for analyses involving four treatment groups than for two treatment groups. For analyses involving only two treatment groups the size of the correlation of the dependent variable with the covariable did not have a systematic effect on the degree of agreement of the generated distributions of the F statistic for analysis of covariance using estimated true scores as the covariable with the theoretical F distribution. However, when using four treatment groups, an increase in the correlation of the dependent variable with the covariable cause a systematic decrease in the agreement of the generated distribution with the theoretical distribution.

The analysis of covariance using estimated true scores for the covariable appeared to be as useful a method for testing hypotheses as Lord's statistic. The generated probabilities of a Type I Error for two-tailed tests were in close agreement with the theoretical probabilities for both test statistics. A slight negative skewness for generated distributions of Lord's statistic caused the probabilities of a Type I Error for one-tailed tests systematically to exceed the probabilities of a Type I Error

for two-tailed tests. Power was essentially the same for both test statistics. The use of estimated true scores was slightly more powerful for the two extreme levels of reliability, .5 and .9, of the covariable, and Lord's statistic was slightly more powerful for the intermediate level, .7.

The results clearly indicated the utility of Lord's statistic and my procedure of ANCOVA using estimated true scores as the covariable, when analyzing data from a quasi experiment where the variables are fallible. They also supported the greater generality of the modified ANCOVA procedure. Thistlethwaite (1969) and Campbell and Erlebacher (1970) provide illustrations of the use of ANCOVA using estimated true scores as the covariable. DeGracie (1968) has more recently proposed a test statistic which he points out is similar to the one that I have proposed and investigated, but which has an asymptotic normal distribution.

As a final note of caution none of the above mentioned analyses provides a completely satisfactory substitution for random assignment of experimental units to levels of the independent variable. Although they provide and test estimates of the treatment effects after controlling for initial between level differences on the covariable, there is no guarantee that the covariable reflects all important initial between levels differences, i.e., all of the above mentioned procedures suffer from the same limitations that apply for ANCOVA on infallible variables. Two excellent references on such limitations are provided by Smith (1957) and Lord (1967). Elashoff (1969) also points out the limitations of ANCOVA when random assignment has not been used in the design as well as several other limitations of ANCOVA. Informally Lee Cronbach, Donald Campbell and I (Campbell and Erlebacher 1970,Errata) have considered the problem of choice among covariates

where the intent is to use the covariate to control for the problem of confounding variables. Unfortunately our thoughts are at too preliminary a stage to be reported here.

Summary

In this paper problems caused by the existence of errors of measurement have been identified for factor analysis, regression analysis, ANOVA, and ANCOVA. At least one detrimental effect was seen to exist for each type of analysis. When a researcher's interest is with infallible variables, he runs the risk of biased results from all of the procedures except ANOVA. The estimates of parameters in all four procedures suffer from inflated error variance. Some partial solutions were indicated, but clearly more work is needed on several of the problems.

Most statistical procedures have been developed for models where variables are assumed to be free from errors of measurement. Since almost all educational research involves use of fallible variables, it is important that the effects of errors of measurement for the various models be understood and that the understanding be reflected in current research practice.

REFERENCES

Adcock, R. J. A problem in least squares. The Analyst, 1878, 5, 53-54.

Allen, R. G. D. The assumptions of linear regression. Economica, 1939, 6, 191-201.

Bartlett, M.S. Fitting a straight line when both variables are subject to error. Biometrics, 1949, 5, 207-212.

Berkson, J. Are there two regressions? J. Amer. Statist. Ass., 1950, 45, 164-180.

Box, G. E. P. The effects of errors in the factor levels and experimental design. Bull. Int. Statist. Inst., 1961, 38, 339-355.

Bryant, N.D. Correcting correlations for restriction in range due to selection on an unmeasured variable. Interim Report, project no. 422001, Grant no. OEG-2-7-070701-4249, Dept. of Health, Ed. and Welfare, U.S. Office of Education, Bureau of Education for the Handicapped, 1970.

Campbell, D. T. and Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Compensatory Education: A national debate. Vol. III of The disadvantaged child, New York: Brunner/Mazel, (in press).

Campbell, D. T. and Stanley, J.C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963, 171-246.

Carroll, J. B. The nature of the data, or how to choose a correlation coefficient. Psychometrika, 1961, 26, 347-372.

Cochran, W. G. Errors of measurement in statistics. Technometrics, 1968, 10, 637-666.

Cochran, W. G. Some effects of errors of measurement on multiple correlation. J. Amer. Statist. Ass., 1970, 65, 22-34.

DeGracie, J. Analysis of covariance when the concomitant variable is measured with error. Unpublished Ph.D. dissertation, Iowa State U., 1968.

Durbin, J. Errors in variables. Rev. Inst. Intern. Statist., 1954, 22, 23-32.

Elashoff, Janet D. Analysis of covariance: a delicate instrument. Amer. Ed. Res. J., 1969, 6, 383-401.

Fisher, R. A. Statistical methods for research workers. (4th ed.) London: Oliver and Boyd, 1932.

Geary, R. C. Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown. Econometrica, 1949, 17, 30-58.

Glass, G. V. Alpha factor analysis of infallible variables. Psychometrika, 1966, 31, 545-561.

Guilford, J. P. Fundamental statistics in Psychology and Education. New York: McGraw Hill, 1959, ch. 14.

Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.

Kaiser, H. F. A second generation little jiffy. Psychometrika, 1970, 35, 401-415.

Lindley, D. V. Regression lines and the linear functional relationship. J. Royal Statis. Soc., B, 1947, 9, 218-244.

Lord, F. M. Large-sample covariance analysis when the control variable is fallible. J. Amer. Statist. Ass., 1960, 55, 307-321.

Lord, F. M. A paradox in the interpretation of group comparisons. Psych. Bull., 1967, 68, 304-305.

Madansky, A. The fitting of straight lines when both variables are subject to error. J. Amer. Statist. Ass., 1959, 54, 173-205.

Neyman, J. Existence of consistent estimates of the direction parameter in a linear structural relation between the variables. Ann. Math. Statist., 1951, 22, 497-512.

Pearson, K. On lines and planes of closest fit to system of points in space. Phil. Mag., 1901, 2, 559-572.

Porter, A. C. The effects of using fallible variables in the analysis of covariance. Unpublished Ph.D. dissertation, University of Wisconsin, 1967.

Porter, A. C. The effects of using fallible variables in the analysis of covariance. A paper presented at the American Educational Research Association meetings, 1968.

Reiersol, O. Confluence analysis by means of instrumental sets of variables. Arkiv. Matematik, Ashonomi: Och Fysik, 1945, 32, 1-119.

Roff, M. The relation between results obtainable with raw and corrected correlation coefficients in multiple factor analysis. Psychometrika, 1937, 2, 35-39.

Roos, C. F. A general invariant criterion of fit for lines and planes where all variates are subject to error. Metron, 1937, 13, 3-20.

Saunders, D. R. Factor analysis I: some effects of chance error. Psychometrika, 1948, 13, 251-257.

Scheffé, H. Fitting straight lines when one variable is controlled. J. Amer. Statist. Ass., 1958, 53, 106-117.

Smith, F. H. Interpretation of adjusted treatment means and regressions in analysis of covariance. Biometrics, 1957, 13, 282-307.

Sutcliffe, J. P. Error of measurement and the sensitivity of a test of significance. Psychometrika, 1958, 23, 9-17.

Thistlethwaite, D. L. Some ecological effects of entering a field of study. J. of Ed. Psych., 1969, 60, 284-293.

Thorndike, R. L. Regression fallacies in the matched groups experiment. Psychometrika, 1942, 7, 85-102.

Tukey, J. W. Components in regression. Biometrics, 1951, 7, 33-69.

Wald, A. The fitting of straight lines if both variables are subject to error. Ann. Math. Statist., 1940, 11, 284-300.